The Report Committee for Xiaoqian Li
Certifies that this is the approved version of the following report:

Structural Equation Modeling of Political Discussion Networks

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

____
Daniel A. Powers

____
Joydeep Ghosh

Structural Equation Modeling of Political Discussion Networks

by

Xiaoqian Li

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

August 2017

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Daniel A. Powers. He has always been there to provide me with invaluable guidance and support. I would like to thank his encouragement, patience, care, and kindness. As both a social scientist and a statistician, he is also a role model to me. I would also like to show my gratitude to my second reader Dr. Joydeep Ghosh. I appreciate that he has allowed me to join his class, which has introduced me to data science, and given me his expertise. I would like to truly thank his support and kindness.

I would like to thank Dr. Stephen G. Walker. I appreciate his patience when answering my questions in his class and his kindness whenever we met. I also want to thank Dr. Mary Parker for her care and support when I took her classes. I would like to thank Dr. Keenan Pituch. He has always answered my questions patiently in his classes. I also want to thank Dr. Tiffany Whittaker, who has offered me great help in her classes. I also want to thank Dr. Matthew Hersh for his kindness and help. I would like to thank Dr. Wenhong Chen, who supports me to join the program and always gives me care, encouragement, and support. I would also like to thank Dr. Joseph Straubhaar, who supports me to join the program and always gives me care and support. I would like to thank Vicki L. Keller as well, who gives me a lot of help and care during my stay in the program.

I would like to express my sincere gratitude to my parents. They are always there, supporting me, encouraging me, and helping me. Without their love and care, I never would make it this far. I truly appreciate their love. I would like to thank all my friends in Austin and beyond as well. Their care and support always accompany me. I truly appreciate their friendship. I would also like to express my gratitude to the Department of Statistics and Data Sciences.

Abstract

Structural Equation Modeling of Political Discussion Networks

Xiaoqian Li, M.S.STAT.

The University of Texas at Austin, 2017

Supervisor: Daniel A. Powers

This study conducts structural equation modeling (SEM) of political discussion networks. It examines multiple relationships between political discussion networks—network size and non-kin composition, political efficacy, and neighborhood conversation. Based on a two-step approach, it first analyzes and revises the measurement model and then analyzes and revises the structural model given the revised measurement model. The proposed SEM model includes ordered categorical variables as factor indicators in the confirmatory analysis and outcome variables in the structural regressions. Traditional estimation and regression methods need to be adjusted accordingly. This study uses WLS estimation and adopts a latent variable approach to study the categorical outcome variables in the SEM. The results show that the hypothesized SEM model is fully supported. Neighborhood conversation positively and directly contributes to political discussion network size as well as the non-kin composition of the networks. It also indirectly affects network size through political efficacy. Political efficacy also has a direct effect on network size.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Studies on social networks have become increasingly popular over the past few decades. Among theses, research on discussion networks has received considerable scholarly attention. One specific type of discussion network is the political discussion network, which refers to networks of social ties among people who discuss politics. Studies on political discussion networks have identified its significant impacts on political behaviors, such as participation in elections, views on elections, and public opinions in general (Huckfeldt & Sprague, 1995; Nickson, 2008). This is due to the fact that political conversations in such networks can provide important information for individuals to make political decisions and allow individuals to assess conflicting ideas (Parker, Parker, & McCann, 2008; Mutz, 2006). Given its importance, research on political discussion networks also studies network characteristics and composition, as well as the antecedents that affect network characteristics and composition (Cowan & Baldassarri, 2017; Eveland, 2009).

This study focuses on studying factors that affect the composition of political discussion networks. Two features of political discussion networks are examined—network size and non-kin composition of political discussants. Network size refers to the number of discussants with whom people discuss political matters. Non-kin composition of political discussants refers to the composition of non-kin ties in the networks. Most research on political discussion has been conducted with core network members, especially kin ties and close friends (Klofstad, McClurg, & Rolfe, 2009). Yet, discussion about politics with non-kin actors is important given that new information and opinions can be obtained. In addition, the larger the political discussion network, the higher the possibility of acquiring new political resources.

The factors that can affect the composition of political networks are political efficacy and neighborhood conversation. Political efficacy refers to the feeling that individual political action can have an impact on the political process (Niemi, Craig, & Mattei, 1991). The development of political efficacy depends largely on people's living environment and life behaviors, especially in the neighborhood (Boardman & Robert, 2000; Ohmer, 2007). The more conversations that people engage in within the neighborhood, the more efficacious their sense of politics are. The development of political efficacy also determines how people participate in political discussions with others. Additionally, neighborhood can be a place where people discuss politics with members outside of their families.

Based on these theories, I hypothesize that both neighborhood conversation and political efficacy can positively affect network size. Neighborhood conversation can positively affect both political efficacy and non-kin composition of political discussion networks. Neighborhood conversation also has an indirect effect on network size through political efficacy. I then propose a structural equation model to analyze the relationships among the variables using the secondary data from the Texas Media and Society Survey. The original sample size is 2,015. As this study only focuses on people who have at least one political discussant and who provided valid answers on all the variables included in this study, the final analytical sample size is 1,592.

Chapter 2: Data and Variables

POLITICAL DISCUSSION NETWORKS

Respondents were asked to list up to three people with whom they discuss government, elections, and politics (Klofstad et al., 2009). They were further asked to answer whether these people were their spouse/partner/significant other, parent, sibling, child, co-worker, neighbor, friend, or other. If their answers were in the other category, they were further asked to specify the roles of their discussants. Based on the two questions, we developed two variables accordingly—*network size* and *non-kin composition*—among people who had at least one political discussant.

The size of political discussion networks refers to the total number of people with whom respondents discuss political matters. For all of the 2,015 respondents who answered the question, about 11.76% of the respondents did not have any political discussant at all; about 4.32% had one political discussant; about 14.49% had two political discussants; about 69.43% had three political discussants. The mean of network size was 2.42 with a standard deviation of 1.02. Previous studies on core networks showed that the mean core network size, including core political discussion networks, was always below three, which is consistent with the statistics in the current study (Klofstad et al., 2009).

For the 1,592 respondents included in our analytical sample, 5.03% had only one political discussant; 15.89% had two political discussants; 79.08% had three political discussants. I created an ordinal variable with two categories—*network size*—coded as 2 for people who had three political discussants (above the mean), which accounted for 79.08%, and coded as 1 for people who had 1 or 2 political discussants (below the mean), which accounted for 20.92% (see Table 2.1).

In addition to network size, this study also investigates the non-kin composition of political discussion networks. I calculated the proportion of non-kin composition of political discussion networks, that is, the number of non-kin discussants divided by the total number of discussants. Non-kin discussants refer to people who were co-workers, neighbors, friends, and any other roles other than kin specified by the respondent. Kin discussants refer to people who where spouses/partners/significant others, parents, siblings, children, and any other roles other than non-kin specified by the respondent. For the analytical sample, the result showed that 22.68% did not have any non-kin ties in their political discussion networks, that is, the non-kin proportion being equal to 0%; 19.10% had a non-kin proportion that was equal to 33.33%; 6.6% had a non-kin proportion that was equal to 50.00%; 26.07% had a non-kin proportion that was equal to 66.67%; 25.57% had a non-kin proportion that was equal to 100%.

I recoded the proportion of discussants who were non-kin into an ordinal variable—*non-kin composition*: non-kin proportion being 0% was coded as 1 (22.68%); non-kin proportion being larger than 0% and less than or equal to 50% was coded as 2 (25.69%); non-kin proportion being larger than 50% and less than 100% was coded as 3 (26.07%); non-kin proportion being 100% was coded as 4 (25.57%) (see Table 2.1).

Political Efficacy

Political efficacy was measured by 4 items about political attitudes based on previous studies (Niemi et al., 1991): "People like me don't have any say about what government does" (Gov); "sometimes politics and government seem so complicated that a person like me cannot really understand what is going on" (Understand); "my vote doesn't matter" (Vote); "I don't know enough to cast an informed vote" (Knowledge). The items formed a 5-point Likert scale from 1 = strongly agree to 5 = strongly disagree, with disagreement being the more "efficacious" answer in each instance (see Table 2.1). The 4 items had a Cronbach's α of 0.82.

Neighborhood Conversation

Neighborhood conversation was measure by 3 items: "How often do you have discussions with other people about things happening in your neighborhood" (Things); "how often do you have discussions with other people about problems in your community" (Problems); "how often do you interact with people in your neighborhood" (Interaction). The items had a 5-point scale from 1 = never to 5 = a few days a week or more often (see Table 1). The 3 items had a Cronbach's α of 0.74.

Sociodemographic and Socioeconomic variables

Sociodemographic and socioeconomic variables included age, gender, race and ethnicity, education, income, household size, employment status (see Table 2.1). Age was a continuous variable ranging from 18 to 93 with a mean of 51.89 and a S.D. of 16.28. Gender was a binary variable with female coded as 1 and male coded as 0. About 51.19%

of the respondents were female and 48.81% were male. Race and ethnicity was a binary variable with White being coded as 1 (65.77%) and non-White being coded as 0 (34.23%). Education was measured by three categories: less than high school or high school education (34.35%), some college (27.58%), bachelor's degree or higher (38.07%). Two binary variables were created accordingly: college and bachelor. Employment status was a binary variable with employed coded as 1 (56.91%) and unemployed coded as 0 (43.09%). Income was treated as a continuous variable (mean = 12.39, S.D. = 4.40) as it had 19 categories from less than $5,000 to $175,000 or more. House size was also treated as a continuous variable ranging from 1 to 12.

Table 2.1: Descriptive Statistics of Variables

N = 1,592	Mean or %	S.D.	Min	Max
Age	51.89	16.28	18	93
Gender				
Male	48.81			
Female	51.19			
Race/ethnicity				
White	65.77			
Non-White	34.23			
Education				
High school or lower (reference group)	34.35			
Some college	27.58			
Bachelor's degree or higher	38.07			
Employment status				
Employed	56.91			
Unemployed	43.09			
Income	12.39	4.40	1	19
Household size	2.73	1.49	1	12
Network size				
1 or 2	20.92			
3	79.08			

Table 2.1 (continued)

Non-kin composition				
Non-king proportion = 0%	22.68			
0% < non-kin proportion ≤ 50%	25.69			
50% < non-kin proportion < 100%	26.07			
Non-kin proportion = 100%	25.57			
Internet political efficacy				
No say	3.22	1.34	1	5
Complex	3.37	1.30	1	5
No vote	3.66	1.33	1	5
No knowledge	3.94	1.19	1	5
Neighbor conversation				
Things happening	2.84	1.37	1	5
Problems	2.50	1.29	1	5
Interaction	3.48	1.39	1	5

Chapter 3: Methodology

STRUCTURAL EQUATION MODELING

Structural equation modeling (SEM) is a collection of statistical techniques that can model the relationships among multiple independent and dependent constructs simultaneously (Kline, 2016). Outcome or dependent variables in SEM are referred to as endogenous variables and every endogenous variable has at least one cause in the model. Causes or independent variables in SEM are referred to as exogenous variables as their causes are unknown in the model. The most general kinds of model in SEM are measurement models and/or structural regression models. Measurement models mainly refer to confirmatory factor analysis (CFA) modeling of the underlying latent variable structure by linking latent variables with their respective observed indicators or variables. Structural regression (SR) models mainly refer to analyses involving multiple regression equations, representing direct or indirect relationships among endogenous and exogenous variables, which can be observed or unobserved (i.e., latent).

The use of SEM offers numerous advantages compared to more standard statistical techniques, such as the analysis of variance and multiple regression (Kline, 2016). First, SEM has the ability to analyze both observed and latent variables. It accounts for measurement error as the error has been estimated and separated from the analysis of relationships among variables and thus lends a more realistic and reliable quality to the analysis. Second, SEM can simultaneously model linkages among multiple exogenous and endogenous variables and estimate both direct and indirect effects. In other words, traditional dependent variables in one regression equation can also be independent variables in another regression equation. Third, SEM offers overall model fit statistics, which allow users to assess whether the model fits data well. The fit statistics include chi-square statistics (χ^2), root mean square error approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), among others.

WLS ESTIMATION

The most common estimation method for SEM is maximum likelihood (ML), which assumes multivariate normality for the joint population distribution of the endogenous variables, given the exogenous variables (Kline, 2016). When categorical variables such as ordinal variables are included together as outcome variables, other estimation strategies are needed. There are four popular strategies to model categorical data: asymptotically distribution-fee (ADF) estimation, Satorra-Bentler scaled χ^2 and standard errors, robust weighted least squares (WLS) estimation methods implementation in the software program Mplus, and bootstrapping (Finney & Distefano, 2006). In this study, I utlize the WLS estimation (Muthén, 1984; Fimnney & Distefano, 2006) in Mplus 7 to handle the issue of involving both ordinal indicators in the CFA model and ordinal outcome variables in the SR model.

Using the WLS estimation provided by Mplus 7, ordinal variables are caputred by their latent response variables, which are assumed to be continous and follow the normal distribution. For instance, if an observed variable X has five ordered categories, four thresholds will be generated, i.e., $\tau_1, \tau_2, \tau_3, \tau_4$. The observed variable X is captured by its latent response variable X^* in the following way.

$$X = \begin{cases} 1, & \text{if } X^* \le \tau_1 \\ 2, & \text{if } \tau_1 < X^* \le \tau_2 \\ 3, & \text{if } \tau_2 < X^* \le \tau_3 \\ 4, & \text{if } \tau_3 < X^* \le \tau_4 \\ 5, & \text{if } X^* > \tau_4 \end{cases} \tag{3.1}$$

Thus, both CFA and SR models concern the latent response variables but not the original observed variables. There are maninly two kinds of WLS estimation, mean-adjusted least sqaures (WLSM) and mean- and variance-adjusted weighted least squares (WLSMV). This study uses WLSMV, which makes adjustment to the estimated degrees of freedom to give more accurate model fit statistics.

HYPOTHESIZED SEM MODEL

This study utilizes SEM to analyze the relationships among neighborhood conversation, political efficacy, political discussion network size, and non-kin composition of political discussion networks. Neighborhood conversation is a latent variable (ξ_1) measured by three 5-point ordinal observed variables (X_1, X_2, X_3). Political efficacy (η_1) is also a latent variable measured by four 5-point ordinal observed variables (Y_1, Y_2, Y_3, Y_4). Political discussion network size is an ordinal variable with two categories (Y_5), which is captured by its latent variable (η_2^*). Non-kin composition of political discussion networks is an ordinal variable with four categories (Y_6), which is captured by its latent variable (η_3^*). This study also controls for several sociodemopgrahic and socioeconomic variables: age (X_4), gender (X_5), race/ethnicity (X_6), college (X_7), bachelor's degree (X_8), employed (X_9), income (X_{10}), and household size (X_{11}). Age, income, and household size are continuous variables. Other control variables are all binary variables.

Based on the theories reviewed before, I generated the hypothesized model, including both the measurement and structural model, using the WLS estimator. In the following, I first presented the equations for both measurement and structural model, including all the variables used in the study. Second, I presented the whole SEM model in a diagram, which only focuses on the relationships between focal endogenous and exogenous variables and does not present any control variables (see Figure 3.1).

The following are the measurement equations:

$$X_1^* = \xi_1 + \delta_1 \tag{3.2}$$

$$X_2^* = \lambda_{X_{21}} \xi_1 + \delta_2 \tag{3.3}$$

$$X_3^* = \lambda_{X_{31}} \xi_1 + \delta_3 \tag{3.4}$$

$$Y_1^* = \eta_1 + \varepsilon_1 \tag{3.5}$$

$$Y_2^* = \lambda_{Y_{21}} \eta_1 + \varepsilon_2 \tag{3.6}$$

$$Y_3^* = \lambda_{Y_{31}} \eta_1 + \varepsilon_3 \tag{3.7}$$

$$Y_4^* = \lambda_{Y_{41}} \eta_1 + \varepsilon_4 \tag{3.8}$$

where X_i^* (i = 1, 2, 3) are latent response variables for the observed variables X_i (i = 1, 2, 3) and Y_j^* (j = 1, 2, 3, 4) are latent response variables for the observed variables Y_j (j = 1, 2, 3, 4). ξ_1 is an exogenous factor. δ_i (i = 1, 2, 3) are error terms for X_i^* (i = 1, 2, 3). η_1 is an endogenous factor. ε_j (j = 1, 2, 3, 4) are error terms for Y_j (j = 1, 2, 3, 4), respectively. The relationships between the observed and latent response variables are represented below:

$$X_i = \begin{cases} 1, & \text{if } X_i^* \le \tau_{X_i 1} \\ 2, & \text{if } \tau_{X_i 1} < X_i^* \le \tau_{X_i 2} \\ 3, & \text{if } \tau_{X_i 2} < X_i^* \le \tau_{X_i 3} \\ 4, & \text{if } \tau_{X_i 3} < X_i^* \le \tau_{X_i 4} \\ 5, & \text{if } X_i^* > \tau_{X_i 4} \end{cases}, \quad i = 1,\ 2,\ 3 \tag{3.9}$$

$$Y_j = \begin{cases} 1, & \text{if } Y_j^* \le \tau_{Y_j 1} \\ 2, & \text{if } \tau_{Y_j 1} < Y_j^* \le \tau_{Y_j 2} \\ 3, & \text{if } \tau_{Y_j 2} < Y_j^* \le \tau_{Y_j 3} \\ 4, & \text{if } \tau_{Y_j 3} < Y_j^* \le \tau_{Y_j 4} \\ 5, & \text{if } Y_j^* > \tau_{Y_j 4} \end{cases}, \quad j = 1,\ 2,\ 3,\ 4 \tag{3.10}$$

where $\tau_{X_i k}$ (i = 1, 2, 3; k = 1, 2, 3, 4) are thresholds for X_i and $\tau_{Y_j k}$ (j = 1, 2, 3, 4; k =1, 2, 3, 4) are thresholds for Y_j.

These measurement equations (3.2) – (3.8) can be expressed in matrix algebra terms as follows:

$$\mathbf{X}^* = \mathbf{\Lambda}_X\, \boldsymbol{\xi} + \boldsymbol{\delta} \tag{3.11}$$

$$\mathbf{Y}^* = \mathbf{\Lambda}_Y\, \boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{3.12}$$

where $\mathbf{X}^*$ is the matrix of the latent response variables for the observed indicators X_i (i = 1, 2, 3). $\mathbf{\Lambda}_X$ is the parameter matrix of pattern coefficients for the X_i indicators. $\boldsymbol{\xi}$ is the matrix of exogenous variables including the latent exogenous factor ξ_1 and the exogenous control variables ξ_i (i = 2, 3, 4, 5, 6, 7, 8, 9) where $X_i \equiv \xi_{i-2}$ (i = 4, 5, 6, 7, 8, 9, 10, 11). **δ** is the matrix of error terms for the latent response variables for the observed indicators X_i (i = 1, 2, 3). $\mathbf{Y}^*$ is the matrix of the latent response variables for the observed indicators Y_j (j = 1, 2, 3, 4). $\mathbf{\Lambda}_Y$ is the parameter matrix of pattern coefficients for the Y_j indicators. **η** is the matrix of the latent endogenous factor η_1 and the latent endogenous variables η_2^* and η_3^* for observed variables Y_5 and Y_6. **ε** is the matrix of error terms for the latent endogenous variables. The matrices are as follows:

$$\mathbf{X}^* = \begin{bmatrix} X_1^* \\ X_2^* \\ X_3^* \end{bmatrix}, \mathbf{\Lambda}_X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{X_{21}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda_{X_{31}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \boldsymbol{\xi} = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \\ \xi_7 \\ \xi_8 \\ \xi_9 \end{bmatrix}, \boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \quad (3.13)$$

$$\mathbf{Y}^* = \begin{bmatrix} Y_1^* \\ Y_2^* \\ Y_3^* \\ Y_4^* \end{bmatrix}, \mathbf{\Lambda}_Y = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{Y_{21}} & 0 & 0 \\ \lambda_{Y_{31}} & 0 & 0 \\ \lambda_{Y_{41}} & 0 & 0 \end{bmatrix}, \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2^* \\ \eta_2^* \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix} \quad (3.14)$$

The following are the structural equations:

$$\eta_1 = \gamma_{11}\,\xi_1 + \gamma_{12}\,\xi_2 + \gamma_{13}\,\xi_3 + \gamma_{14}\,\xi_4 + \gamma_{15}\,\xi_5 + \gamma_{16}\,\xi_6 + \gamma_{17}\,\xi_7 + \gamma_{18}\,\xi_8 + \gamma_{19}\,\xi_9 + \zeta_1 \quad (3.15)$$

$$\eta_2^* = \gamma_{21}\,\xi_1 + \beta_{21}\,\eta_1 + \gamma_{22}\,\xi_2 + \gamma_{23}\,\xi_3 + \gamma_{24}\,\xi_4 + \gamma_{25}\,\xi_5 + \gamma_{26}\,\xi_6 + \gamma_{27}\,\xi_7 + \gamma_{28}\,\xi_8 + \gamma_{29}\,\xi_9 + \zeta_2$$

(3.16)

$$\eta_3^* = \gamma_{31}\,\xi_1 + \gamma_{32}\,\xi_2 + \gamma_{33}\,\xi_3 + \gamma_{34}\,\xi_4 + \gamma_{35}\,\xi_5 + \gamma_{36}\,\xi_6 + \gamma_{37}\,\xi_7 + \gamma_{38}\,\xi_8 + \gamma_{39}\,\xi_9 + \zeta_3 \tag{3.17}$$

where γ_{ij} (i = 1, 2, 3; j = 1, 2, …, 9) are parameters. ξ_i (i = 1, 2, …, 9) are exogenous variables including latent exogenous factor ξ_1 and exogenous control variables ξ_i (i = 2, 3, 4, 5, 6, 7, 8, 9) where $X_i \equiv \xi_{i-2}$ (i = 4, 5, 6, 7, 8, 9, 10, 11). ζ_i (i = 1, 2, 3) are error terms for the latent endogenous factor η_1 and latent response variables η_2^* and η_3^*. The relationships between the observed endogenous variables Y_5 and Y_6 and their respective latent response variables η_2^* and η_3^* are as follows:

$$Y_5 = \begin{Bmatrix} 1, & \text{if } \eta_2^* \le \tau_{Y_5 1} \\ 2, & \text{if } \eta_2^* > \tau_{Y_5 1} \end{Bmatrix} \tag{3.18}$$

$$Y_6 = \begin{Bmatrix} 1, & \text{if } \eta_3^* \le \tau_{Y_6 1} \\ 2, & \text{if } \tau_{Y_6 1} < \eta_3^* \le \tau_{Y_6 2} \\ 3, & \text{if } \tau_{Y_6 2} < \eta_3^* \le \tau_{Y_6 3} \\ 4, & \text{if } \eta_3^* > \tau_{Y_6 3} \end{Bmatrix} \tag{3.19}$$

These structural equations (3.15) – (3.17) can be expressed in matrix algebra terms as follows:

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\,\boldsymbol{\xi} + \mathbf{B}\,\boldsymbol{\eta} + \boldsymbol{\zeta} \tag{3.20}$$

where **Γ** is the parameter matrix for direct effects of exogenous variables on the latent endogenous factor and latent response variables. **B** is the parameter matrix for direct effects of endogenous factors and latent response variables on each other. **ζ** is the matrix for error terms for the latent endogenous factor η_1 and latent response variables η_2^* and η_3^*. The matrices are as follows:

$$\mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} & \gamma_{15} & \gamma_{16} & \gamma_{17} & \gamma_{18} & \gamma_{19} \\ \gamma_{21} & \gamma_{22} & \gamma_{22} & \gamma_{24} & \gamma_{25} & \gamma_{26} & \gamma_{27} & \gamma_{28} & \gamma_{29} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} & \gamma_{35} & \gamma_{36} & \gamma_{37} & \gamma_{38} & \gamma_{39} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \boldsymbol{\zeta} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix} \tag{3.21}$$

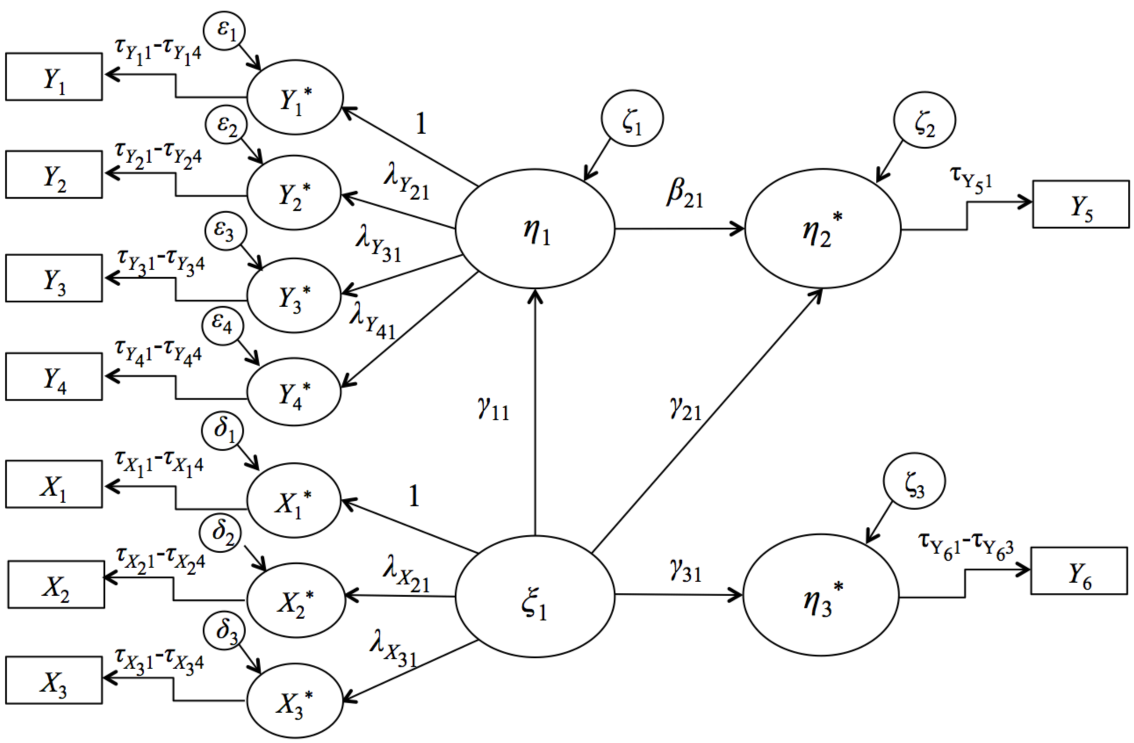


Figure 3.1: Hypothesized SEM Model

Chapter 4: Data Analysis

In the data analysis, I used a two-step modeling approach (Anderson & Gerbing, 1988). In the first step, the measurement model is analyzed by confirmatory factor analysis (CFA) in order to determine whether the measurement fits the data. It focuses on assessing whether observed indicators are well explained by the underlying latent factor. After finalizing the measurement model, the second step is to analyze the structural model. Fit statistics and residuals are checked to finalize the structural model.

MEASUREMENT MODEL

The measurement model was analyzed by confirmatory factor analysis (CFA). The underlying latent factor structure is based on theoretical hypotheses. This study hypothesized that there would be two latent factors in the measurement model—neighborhood conversation (ξ_1) and political efficacy (η_1). Neighborhood conversation was measured by three indicators with a 5-point scale and political efficacy was measured by four indicators with a 5-point scale. As the indicators are ordinal variables, WLSMV estimation was used. The factor loadings concerned the relationship between the latent factor and the latent response variable of each observed indicator. In the initial measurement model, only the two factors were covaried, representing the full structural model, and the errors of the latent response variables are not correlated. Figure 4.1 shows the initial model, which only presents the latent response variables and does not present the observed indicators.

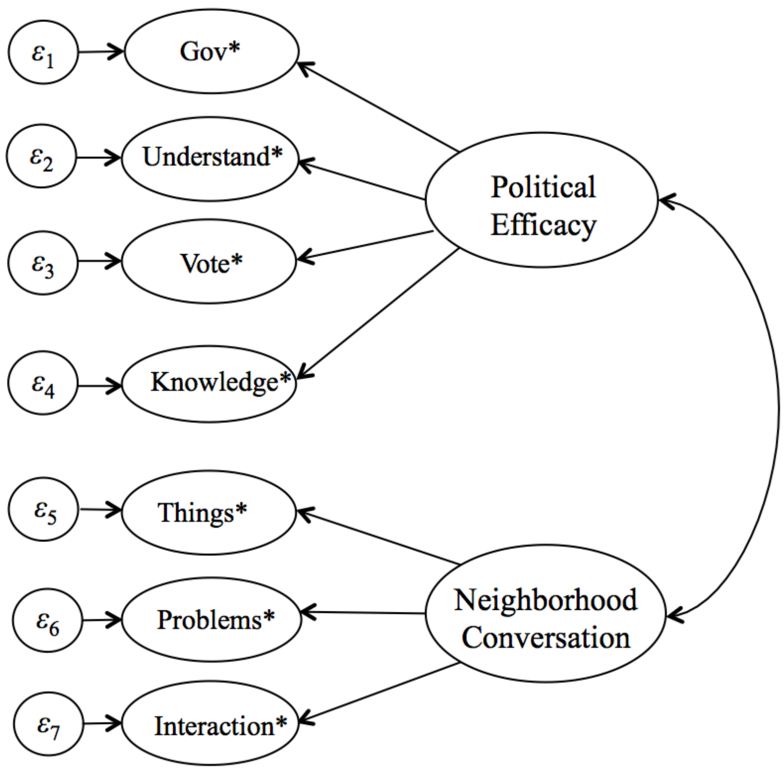


Figure 4.1: Initial Measurement Model

After conducting the CFA for the initial measurement model, I examined the fit indices. The χ^2 test of model fit tested the null hypothesis that the hypothesized model fits the analyzed covariance matrix perfectly. If the χ^2 equals 0, the model perfectly fits the data, which means that each observed covariance equals its predicted counterparts (Kline, 2015). In terms of the initial measurement model, the χ^2 was 366.56 with a degree of freedom of 13 and was statistically significant (χ^2 = 366.560, df = 13, p < 0.001). The initial measurement model rejected the null hypothesis. This suggested that the measurement model did not fit the data well. Yet, χ^2 was sensitive to sample size (Brown, 2015; Kline, 2016). The larger the sample size is, the more likely χ^2 test is significant. Thus, the hypothesis tested by χ^2 is likely to be implausible. Problems with the χ^2 can be solved by relying on other model fit statistics such as RMSEA, CFI, TLI, and so on. The cut-off value of RMSEA indicating a good model fit is around or below 0.05 with a confidence interval captures 0.05 (MacCallum, Browne, & Sugawara, 1996). The cut-off value of CFI or TLI is about or above 0.95 (Marsh, Hau, & Wen, 2004). The RMSEA of the initial measurement model was 0.131 with a 90% confidence interval of 0.119 and 0.142. The CFI was 0.965 and the TLI was 0.944. Thus, the other model fit indices also showed a poor fit of the initial measurement model.

After checking the factor loadings, which were acceptable, and conducting the Lagrange Multiplier test for adding paths, the revised measurement model was obtained and shown in Figure 4.2. The revised measurement model allowed the errors of two latent response variables to be covaried (Gov and Vote). The χ^2 was 39.671 with a degree of freedom of 12 and was still statistically significant (χ^2 = 39.671, df = 13, p < .001). Yet, the RMSEA was 0.038 with a 90% confidence interval of 0.025 and 0.052. The CFI was 0.997 and the TLI was 0.995. Except for χ^2, the other model fit indices were all very good. I also conducted a WLSMV χ^2 difference test. The result showed that $\Delta\chi^2$ was equal to 219.329 with one degree of freedom and was statistically significant ($\Delta\chi^2$ = 219.329, Δdf = 1, p < 0.001). Thus, dropping the error covariance path would result in a significant loss of fit. Thus, both the fit statistics and WLSMV χ^2 difference test supported the revised measurement model. Table 4.1 shows the comparison of the model fit statistics between the initial and revised measurement model.

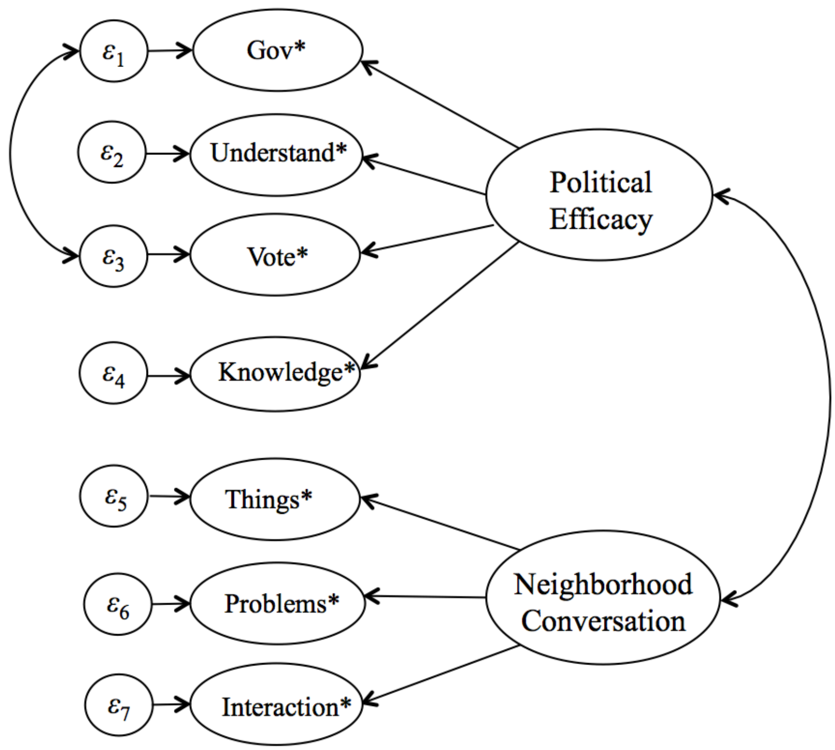


Figure 4.2: Revised Measurement Model

Table 4.1: Initial and Revised Measurement Model Fit Indices

Measurement Model	χ^2	*df*	*p*-value	CFI	TLI	RMSEA	90% C.I. RMSEA
Initial	366.560	13	< 0.001	0.965	0.944	0.131	(0.119, 0.142)
Revised	39.671	12	< 0.001	0.997	0.995	0.038	(0.025, 9.052)

STRUCTURAL MODEL

Given the final measurement model, the initial structural model was conducted. The initial structural model was presented in Figure 4.3. In the initial structural model, all the control variables such as age, gender, race/ethnicity, college, bachelor, employed, income, household size were controlled. The figure only presented the latent factors and latent response variables. The observed variables including factor indicators, observed single indictors, and observed control variables were not presented in the figure.

The χ^2 of the initial structural model was 401.804 with a degree of freedom of 72 and was statistically significant ($\chi^2 = 401.804$, $df = 72$, $p < 0.001$). This poor fit might be due to the large sample size of this data. Yet, the RMSEA was 0.054 with a 90% confidence interval of 0.049 and 0.059. The CFI was 0.967 and the TLI was 0.950. These other model fit indices were good.

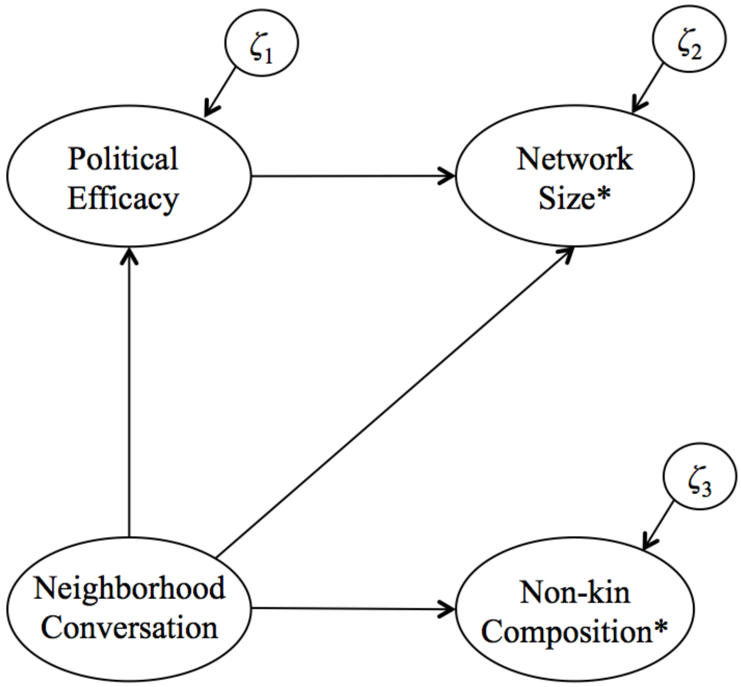


Figure 4.3: Initial Structural Model

After checking the coefficients, which were acceptable, and conducting the Lagrange Multiplier test for adding paths, the revised measurement model was obtained and shown in Figure 4.4. The χ^2 of the revised structural model was 349.683 with a degree of freedom of 71 and was statistically significant (χ^2 = 349.683, df = 71, p < 0.001). This poor fit might be due to the large sample size of this data. Yet, the RMSEA was 0.050 with a 90% confidence interval of 0.045 and 0.055. The CFI was 0.972 and the TLI was 0.958. Except for χ^2, the other model fit indices were all very good. I also conducted a WLSMV χ^2 difference test. The result showed that $\Delta\chi^2$ was equal to 57.989 with one degree of freedom and was statistically significant ($\Delta\chi^2$ = 57.989, Δdf = 1, p <

0.001). Thus, dropping the disturbance covariance path would result in a significant loss of fit. In addition, based on the theories, the larger the political discussion network size is, the more likely the political discussion network contains a non-kin discussant; and vice versa. Thus, the disturbance covariance is acceptable theoretically. Thus, the fit statistics, WLSMV χ^2 difference test, and theories supported the revised structural model. Table 4.2 showed the comparison of the model fit statistics between the initial and revised structural model.

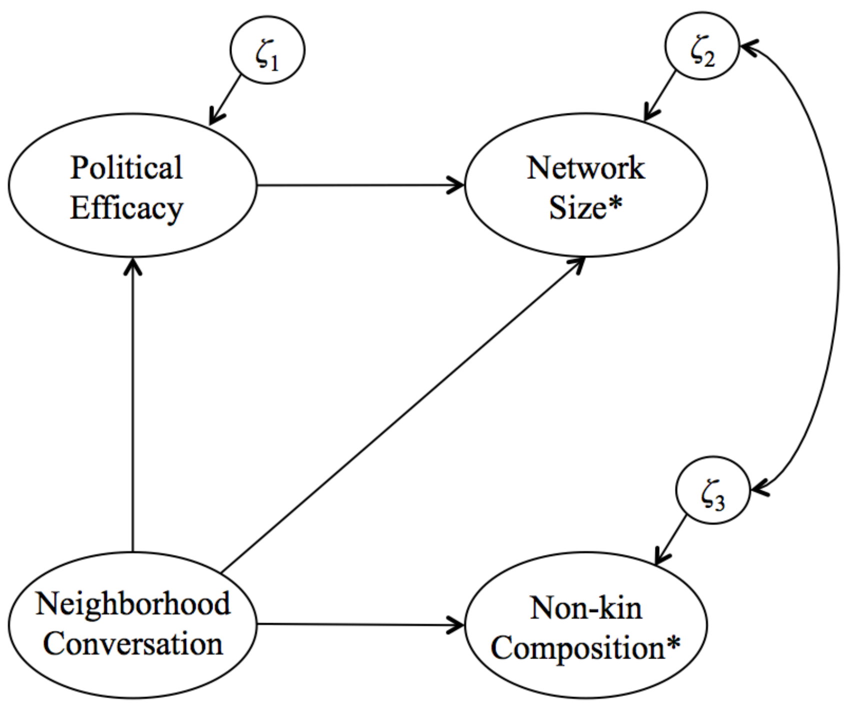


Figure 4.4: Revised Structural Model

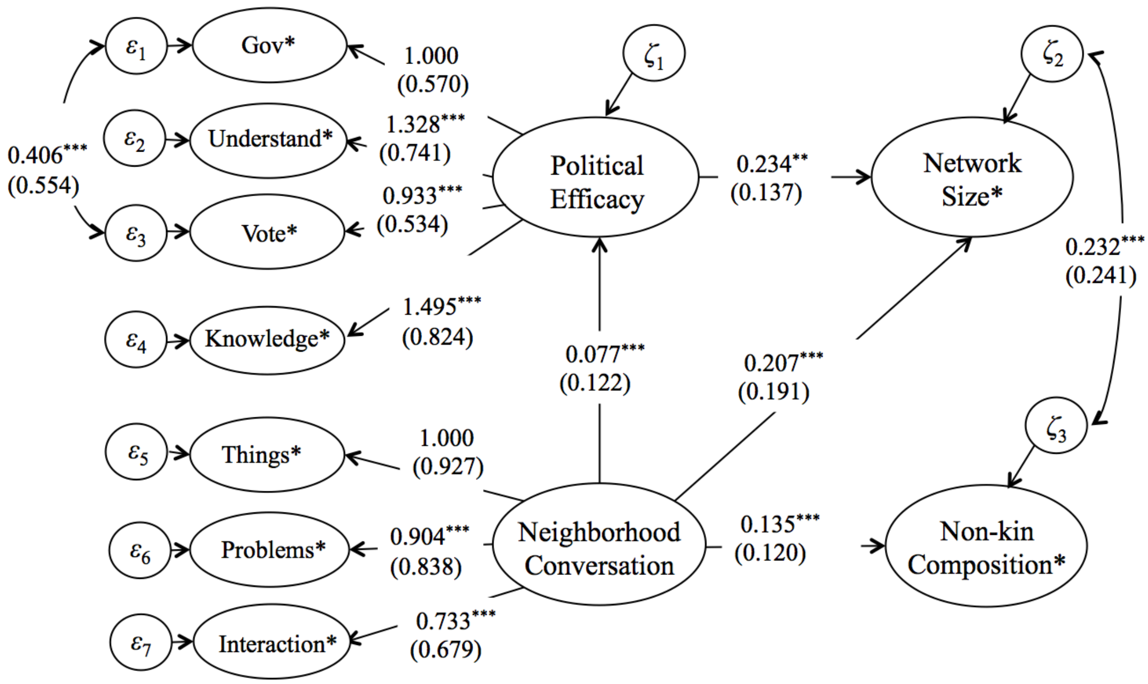
Table 4.2: Initial and Revised Structural Model Fit Indices

Structural Model	χ^2	*df*	*p*-value	CFI	TLI	RMSEA	90% C.I. RMSEA
Initial	401.804	72	< 0.001	0.967	0.950	0.054	(0.049, 0.059)
Revised	349.683	71	< 0.001	0.972	0.958	0.050	(0.045, 0.055)

Chapter 5: Results

The final SEM results were presented in Figure 5.1, controlling for sociodemographic and socioeconomic variables. The model fit indices of the final SEM model were: $\chi^2 = 349.683$, $df = 71$, $p < 0.001$; RMSEA = 0.05, 90% C. I. = (0.045, 0.055); CFI = 0.972; TLI = 0.958. The model fit indices indicated an acceptable model fit. In terms of the measurement part, the unstandardized factor coefficients of Gov, Understand, Vote, and Knowledge for the latent factor political efficacy were 1.000 (fixed in the model), 1.328, 0.933, and 1.495. The standardized factor coefficients of Gov, Understand, Vote, and Knowledge for the latent factor political efficacy were 0.570, 0.741, 0.534, and 0.824, respectively, which were acceptable. The error covariance of Gov and Vote was 0.406. The unstandardized factor coefficients of Things, Problems, and Interaction for the latent factor neighborhood conversation were 1.000 (fixed in the model), 0.904, and 0.733, respectively. The standardized factor coefficients of Things, Problems, and Interaction for the latent factor neighborhood conversation were 0.927, 0.838, and 0.679, respectively, indicating a very good measurement.

The results in the structural part showed that the political discussion network size had a significantly positive association with both political efficacy ($b = 0.234$, S.E. = 0.081, $p < 0.01$) and neighborhood conversation ($b = 0.207$, S.E. = 0.041, $p < 0.001$). Non-kin composition was statistically and positively associated with neighborhood conversation ($b = 0.135$, S.E. = 0.031, $p < 0.001$). Political efficacy was statistically and positively associated with neighborhood conversation ($b = 0.077$, S.E. = 0.019, $p < 0.001$). The error covariance of network size and non-kin composition was 0.232.



Note: The statistics beyond the parentheses are unstandardized coefficients. The statistics in the parentheses are standardized coefficients. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Figure 5.1: Final SEM Model Results

The indirect effects were more specifically examined in the effect decomposition shown in Table 5.1, which also includes the results of sociodemographic and socioeconomic variables having significant indirect effects on political discussion network size through political efficacy. Neighborhood conversation had an indirect effect on political discussion network size through political efficacy (b = 0.018, S.E. = 0.008, $p < 0.05$). Age had an indirect effect on political discussion network size through political efficacy (b = 0.003, S.E. = 0.001, $p < 0.05$). College and bachelor's degree also had an indirect effect on political discussion network size through political efficacy (b = 0.053, S.E. = 0.021, $p < 0.05$; $b = 0.070$, S.E. = 0.026, $p < 0.01$). Income also had an indirect effect on political discussion network size through political efficacy ($b = 0.004$, S.E. = 0.002, $p < 0.05$). All the indirect effects were checked by bootstrapping and the bias-corrected bootstrap 95% confidence intervals showed significant results for all the indirect effects.

Table 5.1: Effect Decomposition of Political Discussion Network Size

Predictor	Direct effect	Indirect effect (via political efficacy)	Total effect
Neighborhood	0.207^{***}	0.018^{*}	0.225
Age	0.000	0.003^{**}	0.003
College	0.033	0.053^{*}	0.086
Bachelor’s degree	–0.036	0.070^{**}	0.034
Income	–0.004	0.004^{*}	0.000

Note: Unstandardized coefficients. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

The results for the control variables were also presented (see Table 5.2). The results showed that age, race/ethnicity, education, and income had a significant effect on political efficacy. People who were older ($b = 0.011$, S.E. = 0.001, $p < 0.001$), non-White ($b = –0.090$, S.E. = 0.036, $p < 0.05$), having a college degree ($b = 0.228$, S.E. = 0.044, $p < 0.005$) or a bachelor’s degree ($b = 0.300$, S.E. = 0.047, $p < 0.001$), and more income ($b = 0.019$, S.E. = 0.005, $p < 0.001$) had a higher level of political efficacy. None of the control variables had any significant relationship with network size. Age, gender, bachelor's degree, income, and household size had a significant relationship with non-kin composition. Those who were older (b = 0.008, S.E. = 0.002, p < 0.001), male (b = –0.281, S.E. = 0.055, p < 0.001), having a bachelor's degree (b = 0.183, S.E. = 0.076, p < .05), employed (b = 0.214, S.E. = 0.063, p < 0.01), having less income (b = –0.034, S.E. = 0.007, p < 0.001) and having a smaller household size (b = –0.085, S.E. = 0.019, p < 0.001) tended to be more likely to have a non-kin tie in their political discussion networks.

Table 5.2: Effects of Control Variables

	Endogenous variables		
	Political efficacy	Network size	Non-kin composition
Control variables			
Age	0.011^{***}	0.000	0.008^{***}
Gender	–0.030	0.098	-0.281^{***}
Race/ethnicity	-0.090^{*}	–0.067	–0.133
College	0.228^{***}	0.033	0.050
Bachelor's degree	0.300^{***}	–0.036	0.183^{*}
Employed	–0.012	0.140	0.214^{**}
Income	0.019^{***}	–0.004	-0.034^{***}
Household size	0.007	–0.016	-0.085^{***}

Note: Unstandardized coefficients. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Chapter 6: Discussion and Conclusion

This study conducts structural equation modeling (SEM) of political discussion networks. Specifically, it examines multiple relationships between political discussion networks—network size and non-kin composition, political efficacy, and neighborhood conversation. The results provide full support of the hypothesized SEM model: neighborhood conversation has a positive direct effect on political discussion network and also has an indirect effect through political efficacy; neighborhood conversation also positively affects non-kin composition of the political discussion networks; political efficacy positively affects network size.

Applying SEM models for categorical variables as outcome variables and mediating variables is more difficult than in the case of linear models. This study has categorical variables as outcome variables and utilizes a WLSMV estimation provided in Mplus 7 to facilitate estimation in this context. Other strategies of handing categorical response variables need to be studied in the future.

References

Boardman, J. D., & Robert, S. A. (2000). Neighborhood socioeconomic status and perceptions of self-efficacy. *Sociological Perspectives*, *43*, 117–136. doi: 10.2307/1389785

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed). New York, NY: Guilford Publications.

Cowan, S. K., & Baldassarri, D. (2017). "It could turn ugly": Selective disclosure of attitudes in political discussion networks. *Social Networks*. Advanced online publication. doi: 10.1016/j.socnet.2017.04.002

Eveland, W. P., & Hively, M. H. (2009). Political discussion frequency, network size, and "heterogeneity" of discussion as predictors of political knowledge and participation. *Journal of Communication*, *59*, 205–224. doi:10.1111/j.1460-2466.2009.01412.x

Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Greenwich, CT: Information Age.

Huckfeldt, R., Beck, P., Dalton, R., & Levine, J. (1995). Political environments cohesive social groups, and the communication of public opinion. *American Journal of Political Science*, *39*, 1025–1054. doi:10.2307/2111668

Kline, R. B. (2016). *Principles and practice of structural equation modeling (*4th ed*)*. New York, NY: Guilford publications.

Klofstad, C. A., McClurg, S. D., & Rolfe, M. (2009). Measurement of political discussion networks: A comparison of two “name generator” procedures. *Public Opinion Quarterly*, *73*, 462–483. doi:10.1093/poq/nfp032

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, *1*, 130–149. doi:10.1037//1082-989x.1.2.130

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320–341. doi: 10.1207/s15328007sem1103_2

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115–132. doi: 10.1007/bf02294210

Mutz, D. C. (2006). Hearing the other side: Deliberative versus participatory democracy. Cambridge, UK: Cambridge University Press.

Nickerson, D. W. (2008). Is voting contagious? Evidence from two field experiments. *American Political Science Review*, *102*, 49–57. doi: 10.1017/s0003055408080039

Niemi, R. G., Craig, S. C., & Mattei, F. (1991). Measuring internal political efficacy in the 1988 National Election Study. *American Political Science Review*, *85*, 1407–1413. doi:10.2307/1963953

Ohmer, M. L. (2007). Citizen participation in neighborhood organizations and its relationship to volunteers' self-and collective efficacy and sense of community.

Social Work Research, *31*, 109–120. doi: 10.1093/swr/31.2.109

Parker, S. L., Parker, G. R., & McCann, J. A. (2008). Opinion taking within friendship networks. *American Journal of Political Science*, *52*, 412–420. doi: j.1540-5907.2008.00320.x